University of Pennsylvania Working Papers in Linguistics

Volume 21
Issue 1 *Proceedings from PLC 38*

Article 10

3-1-2015

How Useful are Transitional Probabilities in Adult-Directed Speech?

Kelly Enochson
kelly.enochson@gmail.com

How Useful are Transitional Probabilities in Adult-Directed Speech?

How useful are transitional probabilities in adult-directed speech?

Kelly Enochson*

1 Introduction

Segmenting a continuous speech stream into discrete words is one of the major tasks of early language learning. One source of readily available information that could be used to accomplish segmentation is transitional probabilities (TPs) between syllables; syllables occurring within a word have high TPs compared to syllables across word boundaries, therefore learners could segment speech by positing potential word boundaries at local TP minima. Laboratory studies investigating statistical learning of language have indeed shown that infants and adults (Saffran, Newport, and Aslin 1996, Aslin, Saffran, and Newport 1998) are able to compute TPs between adjacent syllables and use this information to segment continuous speech into words. Nevertheless there remains disagreement as to whether this is in fact how speech segmentation is accomplished in natural language acquisition.

A number of studies have addressed this issue by incorporating more naturalistic input or adding additional complexity (Johnson and Tyler 2010, Pelucchi, Hay, and Saffran 2009, Thiessen, Hill, and Saffran 2005). However in a series of computational investigations, Yang (2004) and Gambell & Yang (2006) attempted to segment words in a corpus of spontaneous child-directed speech based on TP information alone. What they found was that the abundance of monosyllabic words in child-directed speech made identifying local TP minima effectively useless for postulating word boundaries. The showed instead that eschewing statistical information and instead incorporating primary stress information leads to successful speech segmentation.

Interestingly, there are characteristics of *adult*-directed speech that suggest it may in fact be more amenable to segmentation based on TPs than child-directed speech (Brent and Siskind 2001, Kirchhoff and Schimmel 2005, Ma, Golinkoff, Houston, and Hirsh-Pasek 2011, Soderstrom 2007, Thiessen et al. 2005, Trainor and Desjardins 2002). If it turns out that adult-directed speech provides richer statistical information than child-directed speech in these terms, an important prediction is made: assuming that learners indeed compute syllable-to-syllable TPs as part of the mechanism for postulating word boundaries, then adult-directed speech should be more informative than child-directed speech. Further, adult second language learners should be able to use the type of speech they hear most to segment words in the language from naturalistic (continuous) input.

Here, we explore this issue by using the algorithms developed by Yang and colleagues to determine whether the statistical information available in adult-directed speech is more informative in terms of transitional probability than in child-directed speech. The findings reveal, somewhat surprisingly, that adult-directed speech is not more informative, and further that using TPs as a sole means of speech segmentation remains largely unsuccessful. Following Gambell & Yang (2006), we then show that compared to TPs, stress information allows for significantly more successful segmentation of continuous speech (Johnson and Jusczyk 2001, Johnson and Seidl 2009, Johnson 2008, Shukla, Nespor, and Mehler, 2007, Thiessen and Saffran 2003).

2 Background

In the now classic studies on statistical learning of word boundaries, Saffran et al. (1996) and Aslin et al. (1998) exposed English-speaking adults and infants to unsegmented speech in which "words" were identifiable only by the higher TPs between their syllables. Participants were able to successfully discriminate between words and part-words after a relatively short exposure period. In order to show that segmentation based on TP information can succeed in the face of more naturalistic stimuli, Pelucchi, Hay, & Saffran (2009) exposed 8-month-old infants to fluent, infant-directed Italian speech consisting of repetitions of four 2-syllable words. Results showed that infants were again able to discriminate between words and part-words, suggesting that statistical

*Many thanks to Jennifer Culbertson, Charles Yang, and the audience at PLC 38.

learning mechanisms can scale up to a controlled subset of natural language. However, Johnson & Tyler (2010) argue while infants can use TPs for speech segmentation under simplified conditions, this mechanism may break down when the language becomes more complex. They found that infants as young as 5.5-month-olds were able to segment speech using TPs, but were only successful when word length was uniform across the artificial language. In fact, in a language with both 2- and 3-syllable words, neither 5.5- nor 8-month-olds were able to discriminate words from part-words. These results suggest that while it is *possible* to use available TP information to find word boundaries under some conditions, this may not be sufficient for natural language acquisition.

While both infants and adults are capable of tracking transitional probability in continuous speech, there are other cues to word boundaries available in natural language. Saffran et al. (1996) demonstrated that while distributional cues alone were sufficient, adding a prosodic cue—specifically, final vowel lengthening—facilitated segmentation. Johnson & Jusczyk (2001) and Johnson & Seidl (2009) demonstrated that infants could segment speech based on statistical information alone, but when statistical cues conflicted with prosodic cues, infants favored the latter as a cue to word boundaries. However, prosodic information, unlike statistical information, is language-specific; it is possible, then, that infants rely on prosodic information more heavily at a later stage of acquisition. Indeed, Thiessen & Saffran (2003) demonstrated that while 9-month-old infants favor stress information over statistical information, 7-month-old infants instead favor statistical information.

One way to further investigate whether word segmentation can be done on the basis of TPs, independently of how learners accomplish this task, is to look more closely at the input. Swingley (2005) ran a speech segmentation algorithm on two corpora of infant-directed speech in order to test the robustness of statistical learning as a means of segmenting natural speech. Using a slightly different measure of co-occurrence probability, *mutual information*, Swingley's algorithm correctly identified about 80% of the words in each corpus. Importantly, speech segmentation was modeled using both co-occurrence probability and frequency information. For example, monosyllables that exceeded a certain frequency threshold were automatically identified as words. Bi-syllabic and tri-syllabic sequences had to meet both frequency and mutual information criteria to be identified as words. By contrast, using TP information alone, Yang (2004) and Gambell & Yang (2006) found that only 30% of words were correctly identified in child-directed speech corpora (see below for a more detailed discussion). However, when the algorithm incorporated information about primary stress, it became significantly more successful, correctly identifying about 70% of words from the corpora.

There are a number of differences between adult-direct speech and child-directed speech, with a body of evidence suggesting that many of the characteristics of child-directed speech aid infants in language acquisition. However, a number of the differences between child- and adult-directed speech are prosodic, such as child-directed speech exhibiting greater pitch variation and repetitive intonational structures (Thiessen et al. 2005), which would not impact the usefulness of child-directed speech in terms of statistical information. In fact, speech recognition software trained on child-directed speech did not perform better than software trained on adult-directed speech (Kirchhoff and Schimmel 2005). Additionally, some of the characteristics of child-directed speech, such as higher pitch, specifically do *not* aid in acquisition (Trainor and Desjardins 2002), and must serve some other function such as attracting and keeping an infant's attention.

The current study follows the procedures outlined in Yang (2004) and Gambell & Yang (2006) to determine whether adult-directed speech provides richer statistical information for language learners in terms of transitional probability than child-directed speech. If this proves to be the case, this has implications for both first and second language acquisition. Adult-directed speech may be more informative for infant language learners in terms of transitional probability, providing a means to bootstrap into prosodic information for speech segmentation. Additionally, adult second language learners may be able to use statistical information to segment continuous speech when learning in an immersion environment.

3 Experiment 1

Transitional probability refers to the probability of one syllable occurring given the previous syllable. The formula for TP is given in Figure 1.

$$TP(A \rightarrow B) = \frac{\Pr(AB)}{\Pr(A)}$$

Figure 1: Transitional Probability formula.

A high transitional probability indicates that B is likely given A, while a low transitional probability indicates that B is unlikely given A. Syllables within a word have a higher transitional probability than syllables across word boundaries. For example, in the sequence *pretty baby* the likelihood of the syllable *pre* being followed by the syllable *ty* is quite high, while the probability of the syllable *ty* in *pretty* being followed by the syllable *ba* in *baby* is quite low. Segmenting the two words using TP would result in a word boundary correctly postulated between the sequence *pretty* and the sequence *baby*.

The experiments in this study use transitional probability as the statistic of co-occurrence frequency, and attempt to segment academic adult-directed speech based solely on this measure.

3.1 Method

Data for all three experiments reported here come from Michigan corpus of Academic Spoken English (MICASE) (Simpson, Briggs, Ovens, and Swales 2002). Data were transcribed using the CMU Pronouncing Dictionary (Bartlett, Kondrak, and Cherry 2009), using the Maximize Onset principle. This means that, for instance, the word *Einstein* would be syllabified as *Ein.stein*, using the largest possible English onset, [st]. Primary, secondary, and tertiary stress were all included in the transcription. Words in the corpus not included in the CMU Pronouncing Dictionary were manually transcribed by the author using the CMU Pronouncing Dictionary phoneme set and using standard dictionary pronunciation and stress. Syllable boundaries, word boundaries, and utterance boundaries were all included as delimiters.

The experiments reported here use a learning algorithm that includes two stages: the first is an initial learning stage, in which all data were read and transitional probabilities were computed for all syllable pairs, and the second is a testing stage, in which word boundaries were postulated at points of local minima—a syllable boundary was posited where the transitional probability between two syllables is lower than the TP on either side.

Gambell & Yang (2006) used 226,178 words and 263,660 syllables from infant-directed speech corpora, but found that transitional probability stabilizes after about 100,000 syllables. Based on this, Experiment 1 reported here uses just over 100,000 syllables; data come from 5 study groups in MICASE, including 113,607 words and 137,201 syllables.

3.2 Results

Performance of word segmentation algorithms is typically reported using *precision*, which indicates the proportion of postulated words that are actual words, and *recall*, which indicates the proportion of actual words that are postulated by the algorithm as words. These two measures are weighted and compared using the F-measure, the formula for which is shown in Figure 2. For the purposes of this paper, $\alpha = 0.5$ (i.e., Gambell and Yang 2006, Goldwater, Griffiths, and Johnson 2009, Roark, Mitchell, and Hollingshead 2007).

$$F_a = \frac{1}{a\frac{1}{P} + (1 - a)\frac{1}{R}}$$

Figure 2: F-measure formula.

Results from Experiment 1 suggest that adult-directed speech is not more informative than child-directed speech in terms of transitional probability. Results are shown in Table 1.

	PRECISION	RECALL	F-MEASURE
Yang (2004)	41.6%	23.3%	.299
Experiment 1	37.0%	17.0%	.233

Table 1: Results from Experiment 1 and from Yang (2004).

This means that in Experiment 1, over 60% of items extracted by the program are not words, and more than 80% of the actual words are not extracted. These results are slightly worse than those reported in Yang 2004. Results from Experiment 1 suggest that TP alone is not a viable mechanism for segmentation of the data in this corpus. However, there are possible alternative explanations for the poor performance of the TP algorithm, including the size of the corpus. Experiment 2 doubles the corpus size in an attempt to improve performance.

4 Experiment 2

The poor performance of transitional probability at word segmentation may be due features of the data sample, specifically the higher type/token ratio and the larger vocabulary characteristic of adult-directed speech. Child-directed speech typically includes a reduced number of word types, simplifying the vocabulary as compared to adult-directed speech (Soderstrom 2007). The larger vocabulary of adult-directed speech potentially obfuscates statistical information, and may require larger input to achieve stable transitional probabilities. Experiment 2 doubles the sample size to determine if this is a factor contributing to the low performance in Experiment 1.

4.1 Method

Data for Experiment 2 come from 7 study groups and 2 advising sessions in MICASE, comprising a total of 190,909 words and 228,336 syllables. If the problems with word segmentation occurring in Experiment 1 are the result of adult-directed speech needing more data in order to achieve stable TPs, then doubling the sample size may improve the algorithm's performance.

4.2 Results

Results from Experiment 2 are shown in Table 2.

	PRECISION	RECALL	F-MEASURE
Yang (2004)	41.6%	23.3%	.299
Experiment 1	37.0%	17.0%	.233
Experiment 2	37.6%	17.3%	.237

Table 2: Results from Experiments 1 and 2, and from Yang (2004).

Results from Experiment 2 are only marginally better than the results in Experiment 1, and still lower than the results reported in Yang's studies (Gambell and Yang 2006, Yang 2004). Doubling the sample size did not markedly improve performance of TP at segmenting speech.

Results from the first two experiments indicate that adult-directed speech does not provide richer statistical information for language learners than child-directed speech. As Yang (2004) notes, the primary cause for the failure of speech segmentation algorithms using transitional probability is the preponderance of monosyllabic words, and this proves to be true both in child-directed speech and adult-directed speech. The corpora used here are comprised of data that are 61% monosyllabic words, with a monosyllabic word following another monosyllabic word 77% of the time. By comparison, in Yang's child-directed speech data, a monosyllabic word is followed by another monosyllabic word 85% of the time. Monosyllabic words present a problem for an algorithm that postulates word boundaries at points of local minima—it will be unable to distinguish words with fewer than two or more than three syllables. In the corpora in this paper, 1.8% of words are more than three syllables long, which is also a problem for transitional probabilities using local minima, although clearly not the reason for the algorithm's failure.

Given the failure of statistical learning alone to effectively segment continuous speech, the algorithm used in Experiment 3 makes use of additional information available in the language learner's input. Following Gambell & Yang (2006), Experiment 3 uses stress information along with transitional probability to segment the adult-directed speech corpora used above.

5 Experiment 3

Since statistical learning seems to be ineffective at segmenting continuous speech, Experiment 3 explores what other information in the input can be used in speech segmentation. Recall that Swingley (2005) included an additional parameter that resulted in successful segmentation—in addition to co-occurrence probability information, Swingley's algorithm included a minimum frequency threshold, which specifically aided in segmenting monosyllabic words, thus avoiding the inherent problem with using co-occurrence probability alone. By comparison, Yang (2004) demonstrated that an algorithm using only stress information is quite effective at segmenting infant-directed speech. Gambell & Yang (2006) found that using primary stress information was substantially more effective at segmenting speech because unlike statistical learning, it makes use of single word utterances (Brent and Siskind 2001) and can deal effectively with strings of monosyllabic words. Experiment 3 includes stress information in the segmentation algorithm.

5.1 Method

The algorithm used in Experiment 3 makes use of both statistical information and stress information. Following Gambell & Yang (2006), Experiment 3 includes two parameters for segmenting speech using stress information. First, if two primary stressed syllables are adjacent, a word boundary is postulated between them. Second, if two unstressed syllables are adjacent, a word boundary is postulated at the point of TP local minimum. These parameters allows for segmentation of strings of monosyllabic words, eliminating the major problem with using transitional probability alone.

5.2 Results

Results from Experiment 3 are shown in Table 3.

	PRECISION	RECALL	F-MEASURE
Gambell & Yang (2006)	73.5%	71.2%	.723
Experiment 3	65.9%	77.6%	.713

Table 3: Results from Experiment 3 and from Gambell and Yang (2006).

Using both TP information *and* stress information, the speech segmentation algorithm is quite successful, and on par with the results reported in Gambell & Yang (2006). This suggests that for stress information, as with transitional probability, adult-directed speech does not differ markedly from child-directed speech in terms of richness of information available to the language learner.

6 Discussion

The results here indicate that transitional probability alone is not sufficiently informative to be used as a sole means of segmenting continuous naturalistic speech. Adding information about primary stress significantly improves results by circumventing statistical learning entirely for strings of monosyllabic words, which are the primary problem with TP using local minima. However, this method assumes an innate and domain-specific parameter. In order for primary stress information to successfully segment speech, the listener must come to the problem with the knowledge that words contain only one primary stress. One of the reasons statistical learning has gained widespread popularity as a theory of speech segmentation is that the computation of distributional statistics like TPs can potentially be accomplished by means of domain general cognitive mechanism; tracking and using TPs has been demonstrated in the learning of tone sequences (Creel, Newport, and Aslin 2004, Gebhart, Newport, and Aslin 2009, Saffran, Johnson, Aslin, and Newport 1999), visual feature combinations (Kirkham, Slemmer, and Johnson 2002), and action sequences (Baldwin, Andersson, Saffran, and Meyer 2008, Buchsbaum, Gopnik, Griffiths, and Shafto 2011), as well as in speech segmentation (Hauser, Newport, and Aslin 2001) and non-adjacent dependency learning (Newport, Hauser, Spaepen, and Aslin 2004). The latter two abilities have also been argued to be present in non-human primates. By contrast, including primary stress information in a theory of word segmentation necessitates positing the involvement of a domain-specific mechanism.

The additional information used in Swingley (2005) also presupposes some language-specific knowledge. The minimum frequency threshold above which monosyllables were posited as words was arbitrarily chosen; several different values for this threshold were tested to determine the most informative value. It is unlikely that this threshold is the same from language to language, thus it is unlikely that this parameter is innate.

While primary stress information substantially improves the speech segmentation algorithm used here, it is important to note that this is because monosyllabic words, including some function words (e.g., *of*, *for*, *with*) are given primary stress in the CMU Pronouncing Dictionary. In practice, function words are reduced and not given primary stress. This would impede the effectiveness of stress as a parameter for positing word boundaries.

The corpus chosen for this study is of adult-directed speech, specifically *academic* speech. The hope was to bias the algorithm in favor of working by selecting careful speech about academic topics. Even with this advantage, a segmentation strategy using TP is utterly unable to segment the corpus data because of the number of monosyllabic words. Academic speech still must include function words, most of which are one syllable, and necessarily fail the TP mechanism.

7 Conclusion

This paper investigated whether adult-directed speech might be more informative for language learners than child-direct speech, in terms of transitional probability information potentially used for word segmentation. If this proved to be the case, then adult-directed speech would be informative for both first and second language acquisition: children could make use of overheard adult-direct speech, and second language learners could make use of the type of speech they would likely hear the most in an immersion context. Results indicate that adult-directed speech does not have richer statistical information in these terms, at least at the syllable level. The F-measure increased with an increase in corpus size, but only very slightly, indicating that a larger sample size would not produce more accurate results. One-syllable words, and, to a much lesser extent, more-than-three-syllable words, drive the algorithm's failure, because transitional probability using local minima fails with these words. Speech segmentation that makes use of both transitional probability and primary stress information successfully segmented a corpus of academic adult-directed speech, in line with studies targeting child-directed speech (e.g., Gambell and Yang 2006), suggesting that a domain-specific mechanism must be employed. Gambell & Yang (2006) use stress information in addition to statistical information, while (Swingley 2005) includes additional criteria, including frequency information and an arbitrary percentile threshold, in his speech segmentation program. While these two approaches are quite different, both assume some domain-specific knowledge. Ultimately, adult-directed speech is not markedly different from child-directed speech in terms of transitional probability.

References

Aslin, Richard. N., Jenny R. Saffran, and Elissa L. Newport. 1998. Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science* 9(4):321–324.

Baldwin, Dare, Annika Andersson, Jenny Saffran, and Meredith Meyer. 2008. Segmenting dynamic human action via statistical structure. *Cognition* 106(3):1382–1407.

Bartlett, Susan, Grzegorz Kondrak, and Colin Cherry. 2009. On the syllabification of phonemes. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 308–316. Association for Computational Linguistics.

Bertoncini, Josiane, Ranka Bijeljac-Babic, Peter W. Jusczyk, Lori J. Kennedy, and Jacques Mehler. 1988. An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General* 117(1):21.

Brent, Michael R., and Jeffrey M. Siskind. 2001. The role of exposure to isolated words in early vocabulary development. *Cognition* 81(2):B33–B44.

Buchsbaum, Daphna, Alison Gopnik, Thomas L. Griffiths, and Patrick Shafto. 2011. Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition* 120(3):331–340.

Creel, Sarah C., Elissa L. Newport, and Richard N. Aslin. 2004. Distant melodies: statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(5):1119.

Gambell, Timothy, and Charles Yang. 2006. Word segmentation: Quick but not dirty. Ms., University of Pennsylvania. URL http://www.ling.upenn.edu/~ycharles/papers/quick.pdf

Gebhart, Andrea L., Elissa L. Newport, and Richard N. Aslin. 2009. Statistical learning of adjacent and non-adjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review* 16(3):486–490.

Goldwater, Sharon, Thomas L. Griffiths, and Mark Johnson. 2009. A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition* 112(1):21–54.

Hauser, Marc D., Elissa L. Newport, and Richard N. Aslin. 2001. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78(3):B53–B64.

Johnson, Elizabeth K. 2008. Infants use prosodically conditioned acoustic-phonetic cues to extract words from speech. *The Journal of the Acoustical Society of America* 123(6):EL144–EL148.

Johnson, Elizabeth K., and Peter W. Jusczyk. 2001. Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language* 44(4):548–567.

Johnson, Elizabeth K., and Amanda H. Seidl. 2009. At 11 months, prosody still outranks statistics. *Developmental Science* 12(1):131–141.

Johnson, Elizabeth K., and Michael D. Tyler. 2010. Testing the limits of statistical learning for word segmentation. *Developmental Science* 13(2):339–345.

Jusczyk, Peter W., and Carolyn Derrah. 1987. Representation of speech sounds by young infants. *Developmental Psychology* 23(5):648.

Kirchhoff, Katrin, and Steven Schimmel. 2005. Statistical properties of infant-directed versus adult-directed speech: insights from speech recognition. *The Journal of the Acoustical Society of America* 117(4):2238–2246.

Kirkham, Natasha Z., Jonathan A. Slemmer, and Scott P. Johnson. 2002. Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition* 83(2):B35–B42.

Ma, Weiyi, Roberta M. Golinkoff, Derek M. Houston, and Kathy Hirsh-Pasek. 2011. Word learning in infant-and adult-directed speech. *Language Learning and Development* 7(3):185–201.

Newport, Elissa L., Marc D. Hauser, Geertrui Spaepen, and Richard N. Aslin. 2004. Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology* 49(2):85–117.

Pelucchi, Bruna, Jessica F. Hay, and Jenny R. Saffran. 2009. Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development* 80(3):674–685.

Roark, Brian, Margaret Mitchell, and Kristy Hollingshead. 2007. Syntactic complexity measures for detecting mild cognitive impairment. In *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*, 1–8. Association for Computational Linguistics.

Saffran, Jenny R. 2003. Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science* 12(4):110–114.

Saffran, Jenny R., Elizabeth K. Johnson, Richard N. Aslin, and Elissa L. Newport. 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* 70(1):27–52.

Saffran, Jenny R., Elissa L. Newport, and Richard N. Aslin. 1996. Word segmentation: The role of distributional cues. *Journal of Memory and Language* 35(4):606–621.

Shukla, Mohinish, Marina Nespor, and Jacques Mehler. 2007. An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology* 54(1):1–32.

Simpson, Rita C., Sarah L. Briggs, Janine Ovens, and John M. Swales. 2002. The Michigan corpus of academic spoken English. *Ann Arbor, MI: The Regents of the University of Michigan*.

Soderstrom, Melanie. 2007. Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review* 27(4):501–532.

Swingley, Daniel. 2005. Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology* 50(1):86–132.

Thiessen, Erik D., Emily A. Hill, and Jenny R. Saffran. 2005. Infant-directed speech facilitates word segmentation. *Infancy* 7(1):53–71.

Thiessen, Erik. D., and Jenny R. Saffran. 2003. When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology* 39(4):706.

Trainor, Laurel J., and Renée N. Desjardins. 2002. Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review* 9(2):335–340.

Yang, Charles D. 2004. Universal Grammar, statistics or both? *Trends in Cognitive Sciences* 8(10):451–456.

Linguistics Program, Department of English
George Mason University
Fairfax, Virginia 22030
kenochso@gmu.edu